

Ophuysen, Stefanie van; Behrmann, Lars

## **Die Qualität pädagogischer Diagnostik im Lehrerberuf - Anmerkungen zum Themenheft "Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz"**

*Journal for educational research online 7 (2015) 2, S. 82-98*



Quellenangabe/ Reference:

Ophuysen, Stefanie van; Behrmann, Lars: Die Qualität pädagogischer Diagnostik im Lehrerberuf - Anmerkungen zum Themenheft "Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz" - In: Journal for educational research online 7 (2015) 2, S. 82-98 - URN: urn:nbn:de:0111-pedocs-114911 - DOI: 10.25656/01:11491

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-114911>

<https://doi.org/10.25656/01:11491>

in Kooperation mit / in cooperation with:



**WAXMANN**  
[www.waxmann.com](http://www.waxmann.com)

<http://www.waxmann.com>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### **Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Stefanie van Ophuysen & Lars Behrmann

## **Die Qualität pädagogischer Diagnostik im Lehrerberuf – Anmerkungen zum Themenheft „Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz“**

Angeregt durch Befunde aus internationalen Schulleistungsstudien erfährt die Qualität von Lehrerurteilen in der gegenwärtigen Diskussion zur Lehrerbildung ein besonderes Interesse (Baumert & Kunter, 2006). Die Studien dokumentierten nicht nur unbefriedigende Leistungen der deutschen Schülerschaft (z. B. Artelt, Stanat, Schneider & Schiefele, 2001; Bos et al., 2003), sondern belegten gleichzeitig die geringe Treffsicherheit von Lehrkräften bei der Identifikation von Risikoschülerinnen und -schülern (Artelt et al., 2001). Die Vermutung eines Kausalzusammenhangs zwischen diesen Befunden regte auf bildungsadministrativer Seite dazu an, das Beurteilen als einen zentralen Aufgabenbereich von Lehrkräften festzulegen (Kultusministerkonferenz, 2004) und das Thema Diagnostik stärker in der Lehrerbildung zu verankern (z. B. § 2 LABG1; Terhart, 2000). Die Zentralität der Diagnostik für das professionelle Lehrerhandeln ist dabei insbesondere durch die Annahme begründet, dass sie eine wichtige Voraussetzung für die Unterrichtsgestaltung und die individuelle Förderung ist (Helmke, 2009; Schrader, 2010).

Für die Lehrerbildungsforschung ergab sich in der Folge zunächst die Aufgabe festzustellen, wie gut die lehrerseitige Urteilsqualität in unterschiedlichen Domänen ausfällt und durch welche Merkmale diese beeinflusst wird. Im deutschsprachigen wie im internationalen Raum liegen entsprechend reichhaltige Forschungsbefunde zur Urteilsakkuratheit pädagogischer Diagnosen vor. Diese Studien folgen einem inzwischen fest etablierten methodischen Ansatz. Als Operationalisierung der Urteilsakkuratheit dient ein Abgleich der Lehrerurteile mit den bei Schülerinnen und Schülern erhobenen Merkmalen. Dabei wird gemäß des Komponentenansatzes von Schrader und Helmke (1987) unterschieden, ob das Lehrerurteil im Niveau, der Streuung und/oder der Rangreihung mit den Schülerdaten übereinstimmt. Unterschiede in der methodischen Umsetzung bestehen darin, ob die Lehrkräfte das Zielmerkmal indirekt (z. B. durch die

---

Prof. Dr. Stefanie van Ophuysen (corresponding author) · Dr. Lars Behrmann, Institut für Erziehungswissenschaft, Arbeitsgruppe Forschungsmethoden/empirische Bildungsforschung, Westfälische Wilhelms-Universität Münster, Georgskommende 33, 48143 Münster, Deutschland

E-Mail: [vanophuysen@uni-muenster.de](mailto:vanophuysen@uni-muenster.de)  
[lars.behrmann@uni-muenster.de](mailto:lars.behrmann@uni-muenster.de)

Einschätzung der Merkmalsausprägung auf einer Rating-Skala) oder direkt (z. B. durch die Angabe der erwarteten Punktzahl in einem Testverfahren) einschätzen. Unter den Studien zur Urteilsakkuratheit entfällt der größte Anteil auf Studien zur Rangreihung von Leistungen, während die beiden anderen Komponenten der Urteilsgenauigkeit seltener untersucht werden.

Bezüglich der Rangreihung von Leistungsurteilen kommen Südkamp, Kaiser und Möller (2012) in ihrer Meta-Analyse zu dem Schluss, dass sich Lehrerurteile wie schon bei Hoge und Coladarci (1989) durch eine moderate Urteilsgenauigkeit auszeichnen ( $r = .63$ ). Die Akkuratheit fällt etwas höher aus, wenn nur direkte Urteile betrachtet werden ( $r = .76$ ; indirekte Urteile:  $r = .61$ ). Verglichen mit der Anzahl an Studien, die leistungsbezogene Urteile in den Blick nehmen, ist die Zahl der Studien zur Akkuratheit der Einschätzung nichtkognitiver Merkmale vergleichsweise gering. Jeweils einzelne Studien analysieren die Akkuratheit der Einschätzung von Interesse (Karing, 2009), Lernmotivation (Spinath, 2005), Leistungsängstlichkeit (Urhahne et al., 2010), Stresserleben (Harazd, van Ophuysen, Gerick & Gieske, 2013), Selbstkonzept (Praetorius, Berner, Zeinz, Scheunpflug & Dresel, 2013; Praetorius, Karst, Dickhäuser & Lipowsky, 2011) oder sozialen Beziehungen (van Ophuysen, 2009a, 2009b). Grundsätzlich gilt, dass hier die Urteilsakkuratheit geringer ausfällt als bei Leistungsbeurteilungen (Spinath, 2005).

Während also systematische Unterschiede in der Urteilsgenauigkeit in Abhängigkeit der zu beurteilenden Schülermerkmale aufgezeigt werden konnten, sind die Befunde zu anderen Merkmalen, die als Prädiktoren der Urteilsqualität angenommen wurden, bisher wenig ergiebig. Die Meta-Analyse von Südkamp und Kollegen (2012) zeigt diesbezüglich, dass bislang kaum hinreichend belastbare Befunde vorliegen, die eindeutige Prädiktoren der Urteilsgenauigkeit identifizieren. Lediglich über eine geschickte Wahl der Erhebungsmethoden ist es denkbar, dass sich die Urteilsakkuratheit von Lehrkräften steigern ließe: So sind Urteile generell genauer, je mehr Informationen die Lehrkräfte über das einzuschätzende Merkmal haben. Darüber hinaus verbessern sich die Urteile, wenn sowohl das Lehrerurteil als auch das Ankerkriterium den gleichen Spezifitätsgrad bei der Erhebung aufweisen. Für weitere Urteils- und Aufgabencharakteristika konnten keine weiteren signifikanten Prädiktoren der Urteilsakkuratheit gefunden werden. Insbesondere für den Bereich der Lehrermerkmale lagen nicht genügend Studien vor, um diese im Rahmen der Meta-Analyse als Moderatoren der Urteilsgenauigkeit einbinden zu können.

Ein weiteres, bislang kaum diskutiertes Problem des hier dargestellten Urteilsgenauigkeitsparadigmas liegt darin begründet, dass im schulischen Kontext in der Regel komplexe Urteile über Schülerinnen und Schüler gefällt werden müssen, für die es kein eindeutiges Validitätskriterium gibt. So sind beispielsweise die Fragen, welche Schulnote „korrekt“ oder welche Übergangsempfehlung „richtig“ ist, kaum zu beantworten (für eine detaillierte Diskussion zur Übergangsdagnostik siehe van Ophuysen, 2006b). Eine Einschätzung der Urteilsakkuratheit bei realistischen und schulalltäglichen Diagnosen ist folglich häufig gar nicht sinn-

voll möglich. Für einfache Schülermerkmale, für die es standardisierte, anerkannte Verfahren zur Merkmalerfassung gibt, ist das Untersuchungsparadigma gut anwendbar. Doch spiegelt sich in einem solchen Fall die diagnostische Fähigkeit einer Lehrkraft unseres Erachtens eher darin wider, dass sie dieses adäquate Instrument einsetzt, also darin, dass sie mittels angemessener Verfahren die korrekten Informationen über ihre Schülerinnen und Schüler generiert und nicht darin, dass sie „intuitiv“ den korrekten Wert vorhersagen kann.

Insgesamt wird die für die Lehrerbildungsforschung zentrale Frage, wie die Urteilsqualität von Lehrkräften im Rahmen schulisch relevanter diagnostischer Aufgaben verbessert werden kann, durch diesen Forschungsstrang bislang nicht in befriedigender Weise beantwortet. Die Fokussierung auf die Urteilsgenauigkeit als einzigem Indikator für die Qualität der Diagnostik greift aufgrund fehlender empirischer Belege für die Existenz einer diagnostischen Kompetenz (vgl. Spinath, 2005) sowie aufgrund der Komplexität der zu treffenden Urteile zu kurz und ist durch andere Untersuchungsstrategien zu ergänzen.

Die drei Beiträge im vorliegenden Themenheft nehmen sich dieser Aufgabe an und beleuchten das professionelle Handeln von Lehrkräften auf unterschiedliche Weise. Wir fassen die zentralen Befunde der drei Arbeiten zunächst kurz zusammen. Um die Beiträge mit Blick auf das Thema der Qualität lehrerseitiger Diagnostik einordnen und diskutieren zu können, werden wir im Anschluss daran den Begriff der pädagogischen Diagnostik genauer spezifizieren und ein heuristisches Rahmenmodell vorstellen, das die Ableitung unterschiedlicher Qualitätsaspekte erlaubt.

## **Die Studien des Themenhefts**

In der Arbeit von *Ohle, McElvany, Horz und Ullrich (2015)* werden mit Daten zur Bild-Text-Integration aus dem BiTe-Projekt (Schnotz et al., 2010) die Zusammenhänge verschiedener kognitiver und nichtkognitiver Merkmale von 121 Lehrkräften der Sekundarstufe untersucht, von denen angenommen wird, dass sie für akkurate diagnostische Urteile prädiktiv sind. Zu den kognitiven Merkmalen zählen die Einstellungen zur Diagnostik, die Motivation, diagnostisch zu handeln, sowie auf diagnostisches Handeln bezogene Selbstreflexion und Selbstwirksamkeit. Zu den nichtkognitiven Merkmalen zählen das von Lehrkräften studierte Hauptfach an der Universität sowie das tatsächlich unterrichtete Fach, die Lehrerfahrung sowie die selbstberichteten diagnostischen Aktivitäten im Unterricht. Darüber hinaus wurde für die kognitiven Lehrermerkmale untersucht, mit welcher Faktorenstruktur diese sich am besten abbilden lassen. Es konnte diesbezüglich gezeigt werden, dass ein Vier-Faktoren-Modell die Daten am besten repräsentiert. Weiterhin hing die Lehrerfahrung mit der Motivation zu diagnostischen Handlungen positiv zusammen. Das studierte und das unterrichtete Fach standen jedoch nicht in Zusammenhang mit den kognitiven Merkmalen. Darüber

hinaus zeigte sich, dass es einen Zusammenhang zwischen der für diagnostische Aktivitäten aufgewendeten Zeit und den Einstellungen zur Diagnostik sowie der diagnostikbezogenen Selbstwirksamkeit und Selbstreflexion gab.

In der zweiten Arbeit dieses Themenhefts untersuchen *Schäfer* und *Seidel* (2015), inwiefern bereits Lehramtsstudierende in der Lage sind, Unterrichtseinheiten professionell einzuschätzen. Hierzu ließen die Autorinnen 109 Lehramtsstudierende ein Unterrichtsvideo hinsichtlich der Zielklarheit des Unterrichts sowie bezüglich des Klassenklimas mittels eines quantitativen Fragebogens und einer qualitativen Befragung beurteilen. Diese Einschätzungen wurden anschließend mit den Beurteilungen dreier im Bereich der Unterrichtsbeobachtung erfahrener Wissenschaftlerinnen und Wissenschaftler verglichen. Die Autorinnen konnten zeigen, dass Lehramtsstudierende durchaus relevante Ereignisse des Unterrichts bemerken (*noticing*), häufig jedoch auch auf oberflächliche und irrelevante Ereignisse fokussieren. Die meisten dieser Ereignisse begründeten die Studierenden jedoch eher naiv (*reasoning*), wobei das Lernklima schwerer als die Zielklarheit einzuschätzen war. Generell fand sich eine geringe Übereinstimmung in den Einschätzungen der Studierenden und den Wissenschaftlerinnen und Wissenschaftlern. Hierbei zeigte sich, dass die allermeisten Studierenden durchaus ihre Beurteilungen von Ereignissen des Unterrichts fachlich begründen konnten, jedoch andere Begründungen als die erfahrenen Wissenschaftlerinnen und Wissenschaftlern wählten. Darüber hinaus konnte gezeigt werden, dass die qualitativ und quantitativ erhobenen Daten teilweise in die gleiche Richtung gingen, dieser Zusammenhang mit Korrelationen um .2 jedoch recht niedrig ausfiel.

Schließlich untersuchen *Böhmer*, *Hörstermann*, *Gräsel*, *Krolak-Schwerdt* und *Glock* (2015) in ihrer Arbeit die kognitiven Prozesse, die der Entscheidung von Grundschullehrkräften für eine bestimmte Übergangsempfehlung zugrunde liegen. Hierzu haben 72 erfahrene Grundschullehrkräfte Vignetten über vier verschiedenen leistungsstarke Schulkinder in einem *Mouselab*-Szenario bearbeitet und jeweils eine Übergangsempfehlung ausgesprochen. Anhand dieser Daten wurde dann analysiert, ob Lehrkräfte in Abhängigkeit der Konsistenz der gegebenen Informationen verschiedene Verarbeitungsstrategien anwenden. Es zeigte sich, dass den Lehrkräften generell für eine Übergangsempfehlung Leistungsinformationen wichtiger waren als das Sozialverhalten, welches sich wiederum als relevanter als der soziale Hintergrund herausstellte. Darüber hinaus wurde gefunden, dass Lehrkräfte bei unklaren Schülerfällen insgesamt mehr Informationen abrufen. Die Art der aufgesuchten Information unterschied sich jedoch nicht in Abhängigkeit davon, ob das Schulkind eindeutig oder nur uneindeutig einer Schulform zugeordnet werden konnte. Im Anschluss schätzten die Lehrkräfte ein, wie wichtig ihnen die einzelnen Indikatoren für die Übergangsempfehlung waren. Diese Einschätzungen der subjektiven Relevanz klärten allerdings nur 25 % der Varianz im tatsächlichen Informationsabrufverhalten auf.

## Was ist pädagogische Diagnostik?

Klauer (1982) charakterisiert pädagogische Diagnostik in seiner vielzitierten Definition als „das Ingesamt von Erkenntnisbemühungen im Dienste aktueller pädagogischer Entscheidungen“ (S. 5). Unzweifelhaft gibt es im Alltag von Lehrpersonen Unmengen an zu treffenden Entscheidungen: von der Auswahl des Arbeitsblatts für die nächste Unterrichtseinheit bis hin zur Entscheidung, welches Kind am Förderunterricht teilnimmt oder welche Übergangsempfehlung am Ende der Grundschule erteilt wird. Neben „pädagogisch“ motivierten, auf Optimierung des Lernens ausgerichteten Planungs- oder Handlungsentscheidungen kann aber auch die reine Feststellung und Bewertung pädagogisch relevanter Merkmale bereits als Zielsetzung pädagogischer Diagnostik angesehen werden. Um diese erweiterte Zielsetzung und zusätzlich den Anspruch von Professionalität im Sinne eines reflektierten und an wissenschaftlichen Grundlagen orientierten Handelns zu verdeutlichen, schlagen van Ophuysen, Lintorf und Harazd (2013) folgende Definition vor:

Pädagogische Diagnostik meint das systematische Sammeln von Informationen über einzelne Schüler oder Gruppen von Schülern mittels vornehmlich sozialwissenschaftlicher, wenngleich unstandardisierter Instrumente sowie deren Analyse und Integration, um professionelles pädagogisches Handeln vorzubereiten und zu unterstützen. Im schulischen Kontext sind neben unmittelbaren pädagogischen Entscheidungen auch die Informierung und Berichtlegung sowie die Vergabe von Noten und Qualifikationen als professionelle Handlungsweisen zu fassen. (van Ophuysen, Lintorf & Harazd 2013, S. 190)

Häufig wird auch – wie bei Klauer (1982) – die beiläufige Erfassung von Informationen mittels informeller Methoden bereits als pädagogische Diagnostik verstanden, sofern diese für die Optimierung des pädagogischen Handelns genutzt wird (Ingenkamp & Lissmann, 2008). Ebenfalls gibt es weiter gefasste Definitionen, die neben dem Schüler/der Schülerin auch Aufgaben und Unterrichtsmaterialien oder das unterrichtlichen Handeln als zu diagnostizierende Einheit (Subjekt, Objekt, Prozess) zulassen (vgl. Jäger & Petermann, 1999, S. 11).

Im Bereich der schulpädagogischen Diagnostik macht es Sinn, zwischen Basis- und Förderdiagnostik zu unterscheiden. Unter Basisdiagnostik verstehen wir die regelmäßige Erfassung (und Dokumentation) grundlegender, auf Leistung, Arbeits- und Sozialverhalten bezogener Schülerinformationen. Im schulalltäglichen Kontext spielt diese Basisdiagnostik eine wichtige Rolle, da Lehrkräfte einer Informationspflicht unterliegen und auf Nachfrage Eltern wie Kindern Auskunft über „alle grundsätzlichen Schulangelegenheiten“ geben müssen (z. B. § 44 im Schulgesetz NRW – SchulG). Spätestens bei der Notengebung am Ende des Schulhalbjahres wird auf diese Informationen zurückgegriffen. Wie bei einem „Screening“ gilt es bei der Basisdiagnostik darüber hinaus sicherzustellen, dass

alle Schülerinnen und Schüler den gewünschten (Mindest-)Standard der Gruppe erreicht haben. Erreichen einzelne Schülerinnen und Schüler dieses Ziel nicht (oder übertreffen sie die Erwartungen deutlich), werden sie mittels spezifischer, individualdiagnostischer Verfahren genau unter die Lupe genommen. Die genaue Spezifikation des „Förderbedarfs“ wird somit im Rahmen der Förderdiagnostik ermittelt, um Entscheidungen über die jeweils passenden pädagogischen Maßnahmen treffen zu können.

Der Übergang zwischen Basis- und Förderdiagnostik ist dabei fließend: Je detaillierter die Informationen über einzelne Kinder oder Gruppen von Kindern erhoben werden, desto eher nähert man sich der Förderdiagnostik an. Da es sich bei der Basisdiagnostik ‚nur‘ um eine regelmäßig durchgeführte Absicherungsdiagnostik der Mindeststandards handelt, werden die Informationen hier üblicherweise ohne allzu großen Aufwand erhoben. Bei der Förderdiagnostik sollte in der Regel pro Kind mehr Aufwand betrieben werden.

Generell ist der für die Diagnostik zu betreibende Aufwand bei der Informationsgewinnung abhängig von der Relevanz der Konsequenzen einer pädagogischen Entscheidung: Je relevanter eine Entscheidung für den Entscheidungsempfänger ist, umso höher sind auch die Anforderungen an die Gütekriterien der erhobenen Information. Wenn eine Lehrkraft beurteilen möchte, ob zwei Schulkinder in der nächsten Gruppenarbeitsphase gut zusammenarbeiten werden, dann sind die Auswirkungen ihrer Entscheidungen nur kurzfristig relevant und, indem die Kinder wieder auseinandergesetzt werden, leicht zu korrigieren. Deshalb reichen für eine solche Entscheidung auch Informationen mit verhältnismäßig schwach abgesicherter Evidenz. Hat eine Lehrkraft allerdings eine Schulformübergangsentscheidung zu treffen, muss sie berücksichtigen, dass ihre Empfehlung das Leben des Kindes langfristig und in vielen Bereichen – vom aktuellen Wohlbefinden bis zum späteren beruflichen Erfolg – beeinflussen kann. Somit ist klar, dass die zur Einschätzung herangezogene Information höchstmöglichen Gütekriterien entsprechen sollte.

Offensichtlich ist pädagogische Diagnostik eine Lehreraufgabe, die unterschiedlichen Zielen dient und durch eine breite Palette an konkreten, auf das jeweilige Ziel bezogenen Handlungsweisen gekennzeichnet ist. Diese Vielfalt wird auch in den in diesem Themenheft vorgestellten Studien deutlich.

## **„Pädagogische Diagnostik“ in den Studien des Themenhefts**

Mit Blick auf die obigen Ausführungen zum Begriff pädagogische Diagnostik ist zu konstatieren, dass die drei vorliegenden Arbeiten hier sehr unterschiedlich zu verorten sind. Die jeweils betrachteten diagnostischen Lehreraufgaben unterscheiden sich hinsichtlich ihrer Ziele und Konsequenzen ebenso wie hinsichtlich der zu erfassenden Merkmale und „Merkmalsträger“.



So stellen Böhmer et al. die Erteilung der Übergangsempfehlung am Ende der Grundschulzeit in den Mittelpunkt. Dies ist als klassische diagnostische Aufgabe einzuschätzen, die auf Basis von Informationen über vielfältige Merkmale bei individuellen Schülerinnen und Schülern erfolgt. Sie entspricht noch dazu einer kategorialen Zuweisungsentscheidung und ähnelt so – zumindest bei einer oberflächlichen Betrachtung – dem klassischen Bild von medizinischer oder psychologischer Diagnostik, in der auf Basis einer Reihe von Symptomen ein Krankheitsbild zugeordnet wird (Hammerschmid-Gollwitzer, 1993). Weiterhin ist die auf Basis der diagnostischen Informationen zu treffende Entscheidung von großer Bedeutung, da die Wahl der weiterführenden Schulform mit hoher Wahrscheinlichkeit den Schulabschluss des Kindes bestimmt (Cortina & Trommer, 2003) und seine beruflichen Entwicklungsmöglichkeiten stark beeinflusst (Maaz, Lüdtke, Trautwein & Baumert, 2008). Die Relevanz verstärkend kommt hinzu, dass selbst in Bundesländern mit freiem Elternwillen der Empfehlung der Lehrkräfte in der Regel gefolgt wird (Bos et al., 2003, S. 132).

Mit vergleichsweise geringen nachhaltigen Konsequenzen gehen hingegen Entscheidungen einher, die Ohle et al. als Ziel des von ihnen betrachteten diagnostischen Handelns benennen. Sie fokussieren die Auswahl von Aufgaben und Unterrichtsmaterialien und betrachten dazu die Aufgabenschwierigkeit bezogen auf den spezifischen Bereich der Text-Bild-Integration als zu diagnostizierendes Merkmal. Hier handelt es sich also nicht um eine auf die Schülerinnen und Schüler bezogene diagnostische Aufgabe, sondern um die Beurteilung eines Lernmittels.

Ziel des „diagnostischen“ Handelns in der Untersuchung von Schäfer und Seidel ist schließlich die Bewertung der Unterrichtsqualität auf Basis des Lehrerverhaltens. Dabei werden mit Zielklarheit und Herstellung eines guten Klassenklimas zwei konkrete Bewertungskriterien in den Mittelpunkt gestellt. An dieser Stelle muss ergänzt werden, dass das fokussierte Verhalten – also die Beobachtung des Verhaltens anderer Lehrpersonen und die darauf basierende Bewertung von Unterrichtsqualität – eigentlich nicht als pädagogische Diagnostik bezeichnet werden sollte. Um die medizinische Diagnostik als Vergleich zu nutzen: Wenn eine Ärztin Merkmale des beruflichen Handelns einer Kollegin beobachtet und darauf basierend eine Qualitätsbeurteilung abgibt, wird dies kaum als „Diagnostik“ bezeichnet. Auch im Sinne der recht allgemeinen Definition von Klauer (1982) handelt es sich bei der Sammlung von Informationen über das Lehrerverhalten zur Bewertung von Unterrichtsqualität nicht um eine „Erkenntnisbemühung im Dienste einer aktuellen pädagogischen Entscheidung“. Dies wäre allenfalls dann gegeben, wenn auf Basis der Bewertung Hinweise an die beobachtete Lehrkraft gegeben würden, wie sie ihren Unterricht verändern könnte. Wir schlagen stattdessen vor, diese Art der Informationssammlung und Bewertung, die sich auf den Unterrichtsprozess und seine Bewertung bezieht, als Evaluation zu bezeichnen.



Tabelle 1: Charakteristika der „diagnostischen Aufgabe“ in den drei Studien

	Böhmer et al.	Ohle et al.	Schäfer & Seidel
Zielsetzung	Übergangsempfehlung	Aufgabenauswahl	Bewertung von Unterrichtsqualität
Konsequenz der pädagogischen Entscheidung	Langfristig, bedeutsam, schwer revidierbar	Kurzfristig, eher unbedeutend, leicht revidierbar	Keine pädagogische Entscheidung
„Merkmalsträger“	Individuelle Schülerinnen und Schüler	Aufgaben bzw. Unterrichtsmaterialien	Lehrkräfte
Merkmal	Breites Spektrum an status- und prozessbezogenen Merkmalen, von Eigenschaften und Verhaltensweisen	Sehr enger Fokus: Aufgabenschwierigkeit bezogen auf Text-Bild-Integration	Unterrichtshandeln eingegrenzt auf Zielklarheit und Klassenklima

Die Untersuchungen zeigen, dass mit pädagogischer Diagnostik eine große Bandbreite und Vielfalt an Zielsetzungen verbunden sein kann (vgl. Tabelle 1). Dies sollte im Blick gehalten werden, wenn Fragen nach der diagnostischen Qualität oder gar der diagnostischen Kompetenz von Lehrkräften gestellt werden, denn offensichtlich sind je nach diagnostischer Aufgabe unterschiedliches Wissen und spezifische Fähigkeiten bedeutsam. Nachfolgend stellen wir ein allgemeines Rahmenmodell vor, das auf die Gemeinsamkeiten in dieser Vielfalt fokussiert und somit die Ableitung genereller Qualitätskriterien erlaubt.

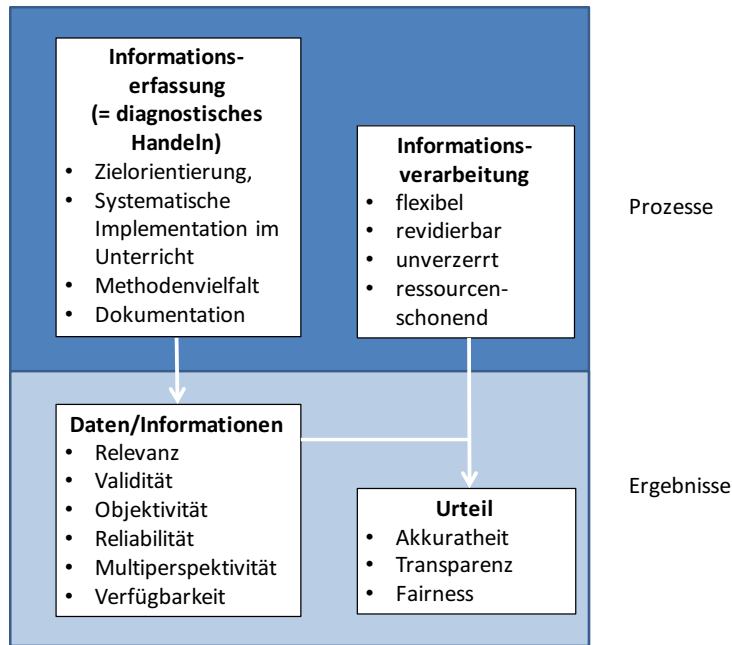
## Kennzeichen „guter“ pädagogischer Diagnostik und die Frage nach der diagnostischen Kompetenz

Diagnostik meint immer einen Prozess, bei dem Informationen erfasst und verarbeitet werden, um zu einem Urteil bzw. zu einer Entscheidung zu kommen. Dies gilt unabhängig von der Zielsetzung, dem Gegenstand der Diagnose oder den zu berücksichtigenden Merkmalen.

Der diagnostische Prozess startet mit der Informationserfassung, die abhängig ist von dem Diagnoseziel, also von der zu treffenden Entscheidung und den zu erfassenden Merkmalen. Es resultieren die für die Diagnose nutzbaren Informationen, die für die kognitive Weiterverarbeitung zur Verfügung stehen. Der kognitive Prozess der Informationsverarbeitung mündet schließlich im diagnostischen Urteil bzw. der pädagogischen Entscheidung.

Mit Blick auf die Diagnosequalität ist also anzunehmen, dass eine hohe Urteilsqualität nur möglich ist, wenn adäquate Verarbeitungsprozesse stattfinden, die auf qualitativ hochwertigen diagnostischen Informationen basieren. Diese wiederum erfordern eine qualitativ hochwertige Informationserfassung. Insgesamt können also mit *Informationserfassung* und *Informationsverarbeitung* zwei Prozesskomponenten und mit *Daten* und *Urteil* zwei Ergebniskomponenten der

Abbildung 1: Heuristisches Rahmenmodell zur Diagnosequalität



Diagnostik unterschieden werden, die in einem spezifischen Abhängigkeitsverhältnis zueinander stehen (siehe Abbildung 1).

Diese systematische Trennung von diagnostischen Komponenten lädt zu Überlegungen ein, welche Qualitätsmerkmale jeweils angelegt werden könnten. Die in Abbildung 1 exemplarisch benannten Merkmale basieren auf theoretischen und normativen Überlegungen. Die postulierten Abhängigkeiten sind in dieser Form (noch) nicht empirisch gestützt.

Wie bereits zu Beginn ausgeführt spielt das Merkmal der Akkuratheit in der bisherigen Forschung zur Bewertung der Urteilsqualität eine zentrale Rolle (Spinath, 2005; Südkamp, Kaiser & Möller, 2012). Transparenz und Fairness sind Kriterien, die bei Entscheidungen mit bedeutsamen Konsequenzen als Maßstab angelegt werden, insbesondere dann, wenn Fragen der sozialen Gerechtigkeit thematisiert werden.

Flexibilität, effiziente Ressourcennutzung sowie Revidierbarkeit sind hingegen Aspekte, die in Untersuchungen betrachtet werden, bei denen die Informationsverarbeitung bei komplexen diagnostischen Aufgaben im Mittelpunkt stehen (Dünnebier, Gräsel & Krolak-Schwerdt, 2009; Krolak-Schwerdt, Böhmer & Gräsel, 2009; Krolak-Schwerdt, Böhmer & Gräsel, 2013; Krolak-Schwerdt & Rummer, 2005; van Ophuysen, 2006a).

Die Merkmale zur Datenqualität lassen sich aus Forderungen an wissenschaftliche Daten ableiten, wenngleich deren Anwendbarkeit im Rahmen der alltäglichen Diagnostik im Schulkontext teilweise in Zweifel gezogen wird (Brookhart,

2003; Gipps, 1994). Jedoch gibt es durchaus auch Vorschläge, wie die klassischen Begrifflichkeiten durch veränderte Operationalisierungen auf den Schulkontext übertragen werden könnten (Brookhart, 2003; Moss, 2003; Smith, 2003).

In der bisherigen Forschung nimmt die Frage nach der Erfassung diagnostischer Informationen recht wenig Raum ein, obwohl dies sowohl grundlegend für den gesamten Prozess ist als auch – da verhaltensbasiert – vergleichsweise hohes Potential für Lernen und Veränderung bieten könnte. Im englischsprachigen Raum wird das diagnostische Handeln unter dem Stichwort *classroom assessment* untersucht. Insbesondere wird erforscht, welche Kriterien (z. B. McMillan, 2001; Randall & Engelhard, 2010) und welche Methoden (z. B. McMillan, 2001; Stiggins & Bridgeford, 1985) Lehrkräfte bei ihren alltäglichen Urteilsaufgaben heranziehen. Während zur Methodennutzung in Deutschland nur vereinzelte Publikationen vorliegen (Edelenbos & Kubanek-German, 2004; Freitag, 2012; van Ophuysen & Lintorf, 2014), wird die gezielte Auswahl diagnostisch relevanter Informationen insbesondere im Kontext der Übergangsempfehlung zunehmend analysiert (Anders, McElvany & Baumert, 2010; Böhmer, Hörstermann, Gräsel & Krolak-Schwerdt, 2012; Riek & van Ophuysen, 2013a, 2013b). Die in Abbildung 1 benannten, möglichen Kriterien einer qualitativ hochwertigen diagnostischen Informationserfassung sind mangels konzeptueller Überlegungen und empirischer Befunde als erste, ad-hoc generierte Vorschläge zu verstehen.

Nach der vielzitierten Konzeptualisierung von Handlungskompetenz nach Baumert und Kunter (2006) kann diese als Konglomerat von verschiedenen Wissensfacetten, speziellen Fähigkeiten sowie Überzeugungen verstanden werden, die für qualitativ hochwertiges Lehrerhandeln förderlich sind. Werden diese Überlegungen mit dem Rahmenmodell zur Qualität pädagogischer Diagnostik gekoppelt, so lassen sich Wissens- und Fähigkeitsbereiche sowie selbst- und handlungsbezogenen Überzeugungen ableiten, die als Komponenten diagnostischer Handlungskompetenz gelten können. Die Erforschung „diagnostischer Kompetenz“ könnte damit die bisherige Output-orientierte Herangehensweise über die Urteilsakkuratheit verlassen und auf die Erfassung von Wissen, Fähigkeiten und Überzeugungen als personenseitige „Input-Faktoren“ ausgeweitet werden. Die Untersuchungen in diesem Themenheft zeigen hier exemplarisch mögliche Fragestellungen und Herangehensweisen auf.

## **„Diagnostische Kompetenz“ in den Studien des Themenhefts**

Die Studie von Böhmer et al. stellt die Auswahl an Informationen für eine komplexe pädagogische Entscheidung in den Fokus. Im schulischen Alltag werden in der Regel für die Übergangsempfehlung keine neuen Informationen erfasst, sondern ein umfassender, über einen mehrjährigen Zeitraum gesammelter Bestand an Informationen über die einzelnen Schülerinnen und Schüler wird zu einer

kategorialen Entscheidung verdichtet. Die Frage, welche der potentiell verfügbaren Informationen dabei berücksichtigt werden, bezieht sich somit auf den Prozess der Informationsverarbeitung. Der verwendete experimentelle Ansatz stellt eine geschickte Methode dar, die gewünschten Informationen verhaltensnah zu erheben und Tendenzen sozialer Erwünschtheit zu vermeiden, die bei direkten, verbal erfragten Relevanzeinschätzungen verzerrend wirken könnten.

Die Studie macht zunächst deutlich, dass das Auswahlverhalten von der Eindeutigkeit und Konsistenz der Informationen selbst abhängig ist. Informationen, die zu einem facettenreichen Bild des zu beurteilenden Schulkinds führen und nicht eindeutig für eine Schulform sprechen, erfordern die Berücksichtigung von umfangreicheren Informationen. Diagnostische Kompetenz ist in diesem Sinne daran zu erkennen, dass diese Widersprüchlichkeit wahrgenommen und dass ihr mit erweiterter Informationssuche begegnet wird.

Weiterhin untersuchen die Autorinnen und Autoren die Bedeutsamkeit verschiedener inhaltlicher Bereiche für die Übergangsempfehlung. Diesbezüglich konnten keine Unterschiede in Abhängigkeit von der Konsistenz der Schülerfälle nachgewiesen werden. Die Reihenfolge und Häufung der abgerufenen Informationen spiegelte insbesondere auch die rechtlichen Vorgaben wider, nach denen die Empfehlung primär auf Leistungsstand und Lernentwicklung rekurrieren nicht aber den sozialen Hintergrund berücksichtigen soll. Es ist hierbei jedoch einschränkend zu berücksichtigen, dass die Befunde über die untersuchten Entscheidungsprozesse aufgrund der artifiziellen Untersuchungssituation möglicherweise nicht eins zu eins in die Praxis übertragbar sind. So könnten die Lehrkräfte beispielsweise die Übergangsentscheidung für nicht real existierende Kinder als weniger relevant beurteilen, was Auswirkungen auf die Qualität des Urteilsprozesses haben könnte.

Nichtsdestotrotz kann das Wissen über die für eine Entscheidung relevanten Urteilkriterien generell als Facette diagnostischer Handlungskompetenz verstanden werden. Problematisch ist jedoch, dass für viele komplexere pädagogische Entscheidungen keine hinreichende wissenschaftliche Evidenz besteht, welches die jeweils relevanten Aspekte sind. Somit ist die Bewertung der diagnostischen Kompetenz auf Basis der für eine Entscheidung berücksichtigten Inhaltskategorien in vielen diagnostischen Situationen bislang nicht möglich. Umso wichtiger ist es zukünftig also, dass in laborexperimentellen Arbeiten identifizierte Prädiktoren diagnostischer Entscheidungen auch im Feld auf ihre Validität hin untersucht werden.

In eine ähnliche Richtung wie die Studie von Böhmer et al. weist die Arbeit von Schäfer und Seidel. Obwohl sie nach unserer Meinung keinen diagnostischen Prozess im engeren Sinne betrachten, existieren offensichtliche Ähnlichkeiten, die es erlauben, Übertragungen auf das Feld der Diagnostik vorzunehmen. So verstehen die Autorinnen unter dem Begriff des „professionellen Blicks“ (*professional vision*) die korrekte Wahrnehmung relevanter Unterrichtsmerkmale (*noticing*) und deren Bewertung unter Bezugnahme auf theoretisch fundiertes, pädagogisches Handlungswissen (*reasoning*). Aus dem verfügbaren Strom

von Informationen gilt es, diejenigen herauszufiltern, die für eine Einordnung und Bewertung des Unterrichts bedeutsam sind. Die Autorinnen arbeiten heraus, dass dieser Filterprozess durch Vorwissen im Sinne von wissenschaftlichen oder vorwissenschaftlichen Theorien beeinflusst wird. Sie bewerten die Deutung der Wahrnehmung als umso kompetenter, je stärker diese durch pädagogisches Fachwissen an Stelle von subjektiven Alltagstheorien geprägt ist, d. h., je dichter die Interpretationen der Lehramtsstudierenden an denen der Expertinnen und Experten liegen. Hier ist jedoch anzumerken, dass nicht per se davon ausgegangen werden kann, dass Lehrkräfte – auch mit großer beruflicher Erfahrung – ein ähnliches Niveau in der fachlichen Einschätzung von Unterrichtsvideos erreichen, wie in der Unterrichtsbeobachtung erfahrene Forscherinnen und Forscher. Die resultierenden Unterschiede zwischen Studierenden und Wissenschaftlerinnen und Wissenschaftler sind somit möglicherweise nicht vollständig auf die fehlende Berufserfahrung der Studierenden zurückzuführen.

Wie in der Studie von Böhmer et al. geht es auch hier um die Frage nach den *relevanten* Informationen, die erfasst und verarbeitet (also zueinander in Beziehung gesetzt, interpretiert und gedeutet) werden müssen. Die Verzahnung von Diagnostik mit didaktischem oder pädagogischem Wissen wird erneut deutlich: Eine gelungene Diagnose startet mit der begründeten Auswahl der relevanten Inhalte. Die Crux für die Lehrkräfte – und für Wissenschaftlerinnen und Wissenschaftler, die sich mit diagnostischer Kompetenz auseinandersetzen möchten – ist an dieser Stelle der Mangel an fundierten wissenschaftlichen Erkenntnissen über bedeutsame Kriterien. Somit bleiben die Studien, die sich der Frage der Kriterienrelevanz annehmen, bislang auch zurückhaltend mit der Verwendung des Kompetenzbegriffs.

Der Kompetenzbegriff ist hingegen in der Studie von Ohle et al. prominent. Sie rekurrieren auf das Modell von Baumert und Kunter (2006) und bezeichnen die von ihnen gemessenen Konstrukte (Motivation, diagnostisch zu handeln, Einstellung gegenüber diagnostischem Verhalten, auf Diagnostik bezogene Selbstreflexion und Selbstwirksamkeit) als Facetten diagnostischer Handlungskompetenz. Sie berichten einen positiven Zusammenhang dieser Kompetenzfacetten mit der tatsächlich gezeigten „diagnostischen Aktivität“. Hier sollte angemerkt werden, dass die gewählten Operationalisierungen im Hinblick auf die inhaltliche Vielfältigkeit der erhobenen Konstrukte kritisierbar sind: Beispielsweise ist die Tatsache, dass sich eine Lehrkraft Gedanken über die Fähigkeiten ihrer Schülerinnen und Schüler und die Schwierigkeit von Aufgaben macht, unseres Erachtens noch kein hinreichender Beleg für tatsächliche diagnostische Handlungen im Unterricht (vgl. Definition der pädagogischen Diagnostik, dieser Beitrag). Davon abgesehen erinnern die im Beitrag von Ohle et al. verwendeten Begrifflichkeiten an die Theorie des geplanten Verhaltens (Ajzen, 1991), in der eine verhaltensprädiktive Intention unter anderem vorhergesagt wird durch verhaltensbezogene Kosten-/Nutzenerwartungen sowie Kontrollüberzeugungen. Nach dieser Theorie erhöhen positive Überzeugungen gegenüber Diagnostik also die Wahrscheinlichkeit, dass überhaupt diagnostisch gehandelt wird – eine notwendi-

ge wenngleich nicht hinreichende Bedingung für qualitativ hochwertiges diagnostisches Handeln. Trotz kritisierbarer Operationalisierungen und der Fokussierung auf einen sehr engen, wenig typischen Diagnosekontext, liefert die Arbeit dennoch Anregungen, sich mit dem Thema von verhaltens- und selbstbezogenen Überzeugungen und ihren Konsequenzen für das diagnostische Verhalten tiefergehend auseinanderzusetzen.

## Ausblick

Die vorliegenden Studien thematisieren die Rolle pädagogischen Wissens und diagnostikbezogener Einstellungen/Überzeugungen für qualitativ hochwertiges, diagnostisches Handeln. Im Sinne von Baumert und Kunter (2006) erscheint es naheliegend, diese Komponenten als Facetten diagnostischer Kompetenz von Lehrkräften zu bezeichnen, wenngleich es aufgrund fehlender empirischer Belege für ihre Existenz einer angepassten Definition der diagnostischen Kompetenz bedarf.

Die Frage nach den relevanten Diagnoseinhalten ist extrem bedeutsam, da ein noch so laborierter Diagnoseprozess, der sich auf die falschen Inhalte bezieht, letztlich eine pädagogische Entscheidung nicht sinnvoll unterstützen kann. Fast jedwede anwendungsorientierte Forschung, die sich auf schulpädagogisches oder fachdidaktisches Lehrerhandeln bezieht, kann somit erkenntnisreich für die pädagogische Diagnostik sein.

Doch andererseits erschweren auch methodisch schlecht erfasste Informationen oder Verzerrungen in der Informationsverarbeitung die Ableitung pädagogischer Entscheidungen. Gerade für pädagogische Entscheidungen mit bedeutsamen und nachhaltigen Konsequenzen sind qualitativ hochwertige Daten erforderlich. Bisherige Studien deuten an, dass in der Schule viele Informationen informell und beiläufig erfasst und teilweise nicht systematisch dokumentiert werden, so dass die Datenqualität für spätere Entscheidungen nicht optimal ist (Edelenbos & Kubanek-German, 2004; Freitag, 2012). Doch bislang ist die Frage, wie Lehrkräfte an Schulen (in Deutschland) ihren diagnostischen Aufgaben nachkommen, kaum empirisch beleuchtet. Entsprechende Untersuchungen könnten aufzeigen, an welchen Stellen Lehrerverhalten verbesserungswürdig ist.

Aufbauend auf solchen Studien zur Informationserfassung und zusammen mit den bisherigen Studien zur Informationsverarbeitung, in denen theoriegeleitet Phänomene der pädagogischen Diagnostik beschrieben werden, könnten dann Interventionsstudien etabliert werden, um Möglichkeiten zur gezielten Verbesserung der diagnostischen Qualität zu eruieren.

Insgesamt gibt es im Bereich der pädagogischen Diagnostik weiterhin einen hohen Forschungsbedarf. Es ist erfreulich zu sehen, dass das Thema zunehmend breiter aufgefächert wird, und deutlich über Studien zur Urteilsakkuratheit hinausgeht. Der Einsatz vielfältiger und innovativer Untersuchungsansätze

(*Mouselab*; Videostudie) und die Kombination experimenteller und nichtexperimenteller Designs in Labor und Feld sind hilfreich, um sich dem Thema in seiner Breite zu nähern. Gleichzeitig erscheint es sinnvoll, Begrifflichkeiten und Operationalisierungen zu schärfen, damit das Thema nicht als beliebig und kon-turlos wahrgenommen wird und gegenüber anderen, eng mit der Diagnostik ver-zahnten Bereichen des Lehrerhandelns, abgrenzbar bleibt. Die Etablierung eines übergreifenden Rahmenmodells zum diagnostischen Handeln – für das in diesem Beitrag ein erster Vorschlag gemacht wurde – könnte hier hilfreich sein. Nicht nur im empirischen, sondern auch im Bereich der theoretischen Konzeptualisierung ist somit noch Handlungsbedarf erkennbar.

## Danksagung

Wir danken Frau Dr. Katrin Lintorf, Bergische Universität Wuppertal, für anregen-de Diskussionen und wertvolle Hinweise im Vorfeld dieser Veröffentlichung.

## Literatur

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- Anders, Y., McElvany, N. & Baumert, J. (2010). Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule. Wie differenziert urteilen Lehrkräfte? In K. Maaz, J. Baumert, C. Gresch & N. McElvany (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (S. 313–330). Bonn: Bundesministerium für Bildung und Forschung.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Test-konzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske+Budrich.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Böhmer, I., Hörstermann, T., Gräsel, C. & Krolak-Schwerdt, S. (2012). *Die Informationsuche von erfahrenen und angehenden Lehrkräften bei der Übergangsempfehlung zur weiterführenden Schule*. Vortrag auf dem 48. Kongress der Deutschen Gesellschaft für Psychologie, Bielefeld.
- Böhmer, I., Hörstermann, T., Gräsel, C., Krolak-Schwerdt, S. & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? *Journal for Educational Research Online*, 7(2), 59–81.
- Bos, W., Lankes, E.-M., Schwippert, K., Valtin, R., Voss, A., Badel, I. & Plabmeier, N. (2003). Lesekompetenzen deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 69–142). Münster: Waxmann.



- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Cortina, K. S. & Trommer, L. (2003). Bildungswege und Bildungsbiographien in der Sekundarstufe I. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland* (S. 342–391). Reinbek: Rowohlt Taschenbuch.
- Dünnebier, K., Gräsel, C. & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. Eine experimentelle Studie zu Ankereffekten. *Zeitschrift für Pädagogische Psychologie*, 23(3-4), 187–195.
- Edelenbos, P. & Kubanek-German, A. (2004). Teacher assessment: the concept of 'diagnostic competence'. *Language Testing*, 21(3), 259–283.
- Freitag, S. (2012). *Pädagogisch-diagnostisches Handeln im Lehrerberuf – eine explorative Fallanalyse*. Unveröffentlichte Masterarbeit: Westfälische Wilhelms-Universität, Münster.
- Gesetz über die Ausbildung für Lehrämter an öffentlichen Schulen (Lehrerausbildungsgesetz – LABG) in der Fassung vom 12.05.2009 (GV. NRW. S. 308), zuletzt geändert durch Gesetz vom 28.05.2013 (GV. NRW. S. 272).
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.
- Hammerschmid-Gollwitzer. (1993). *Wörterbuch medizinischer Fachausdrücke*. München: Wilhelm Goldmann.
- Harazd, B., van Ophuysen, S., Gerick, J. & Gieske, M. (2013). How accurately do principals evaluate the strain experience of their teaching staff? *Journal for Educational Research Online*, 4(2), 65–83.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch Pädagogische Diagnostik*. Weinheim: Beltz.
- Jäger, R. S. & Petermann, F. (1999). Einleitung. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik* (S. 11–13). Weinheim: Psychologie Verlags Union.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(3-4), 197–209.
- Klauer, K. J. (1982). Perspektiven Pädagogischer Diagnostik. In K. J. Klauer (Hrsg.), *Handbuch der pädagogischen Diagnostik* (Bd. 1, S. 3–14). Düsseldorf: Schwann.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland]. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften (Beschluss der Kultusministerkonferenz vom 16.12.2004)*. Bonn.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. *Zeitschrift für Pädagogische Psychologie*, 23(3-4), 175–186.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: a social cognitive analysis. *Social Psychology of Education*, 16(2), 215–239.
- Krolak-Schwerdt, S. & Rummer, R. (2005). Der Einfluss von Expertise auf den Prozess der schulischen Leistungsbeurteilung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(4), 205–213.
- Maaz, K., Trautwein, U., Lüdtke, O. & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, 2(2), 99–106.

- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20–32.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13–25.
- Ohle, A., McElvany, N., Horz, H. & Ullrich, M. (2015). Text-picture integration – Teachers' attitudes, motivation and self-related cognitions in diagnostics. *Journal for Educational Research Online*, 7(2), 11–33.
- Praetorius, A. K., Berner, V. D., Zeinz, H., Scheunpflug, A. & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *The Journal of Educational Research*, 106(1), 64–76.
- Praetorius, A. K., Karst, K., Dickhäuser, O. & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 58(2), 81.
- Randall, J. & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–1380.
- Riek, K. & van Ophuysen, S. (2013a). Kriterien der Übergangsempfehlung – Eine qualitative Interviewstudie mit Grundschullehrkräften. In B. Kopp, S. Martschinke, M. Munser-Kiefer, M. Haider, E. Kirschhock, G. Ranger & G. Renner (Hrsg.), *Individuelle Förderung und Lernen in der Gemeinschaft. Jahrbuch Grundschulforschung* (Bd. 17, S. 270–273). Wiesbaden: VS.
- Riek, K. & van Ophuysen, S. (2013b). *Die Erfassung der Relevanz der Kriterien für die Übergangsempfehlung: Ein Vergleich unterschiedlicher methodischer Zugänge*. Vortrag auf der Jahrestagung der DGfE-Kommission Grundschulforschung, Braunschweig.
- Schäfer, S. & Seidel, T. (2015). Noticing and reasoning of teaching and learning components by pre-service teachers. *Journal for Educational Research Online*, 7(2), 34–58.
- Schnotz, W., Horz, H., McElvany, N., Schroeder, S., Ullrich, M., Baumert, J., Hachfeld, A. & Richter, T. (2010). Das BITE-Projekt: Integrative Verarbeitung von Bildern und Texten in der Sekundarstufe. Projekt BITE. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 143–153). Weinheim: Beltz.
- Schrader, F.-W. (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D.-H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 102–108). Weinheim: PVU.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern. Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.
- Schulgesetz für das Land Nordrhein-Westfalen (Schulgesetz NRW – SchulG) vom 15.02.2005 (GV. NRW. S. 102) zuletzt geändert durch Gesetz vom 25.06.2015 (GV. NRW. S. 499).
- Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26–33.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1–2), 85–95.
- Stiggins, R. J. & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271–286.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762.
- Terhart, E. (2000). *Perspektiven der Lehrerbildung in Deutschland. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Kommission*. Weinheim: Beltz.

- Urhahne, D., Zhou, J., Stobbe, M., Chao, S.-H., Zhu, M. & Shi, J. (2010). Motivationale und affektive Merkmale unterschätzter Schüler: Ein Beitrag zur diagnostischen Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 24(3–4), 275–288.
- van Ophuysen, S. (2006a). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnempfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38(4), 154–161.
- van Ophuysen, S. (2006b). Zur Problematik der Schulformempfehlung nach der Grundschulzeit und ihrer prognostischen Qualität. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 14, S. 49–79). Weinheim: Juventa.
- van Ophuysen, S. (2009a). Die Einschätzung sozialer Beziehungen der Schüler nach dem Grundschulübergang durch den Klassenlehrer. *Unterrichtswissenschaft*, 37(4), 330–346.
- van Ophuysen, S. (2009b). „Meine Schüler verstehen sich gut miteinander“. *Wie genau beurteilen Klassenlehrer das soziale Klima ihrer Klasse?* Vortrag auf der 73. Tagung der Arbeitsgruppe für empirische pädagogische Forschung (AEPF), Bochum.
- van Ophuysen, S., Lintorf, K. & Harazd, B. (2013). Zur Qualität professioneller pädagogischer Diagnostik im Schulalltag – Forschungsbefunde und -desiderate. In K. Schwippert, M. Bonsen & N. Berkemeyer (Hrsg.), *Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven* (S. 187–202). Münster: Waxmann.
- van Ophuysen, S. & Lintorf, K. (2014). Unterschiede in der diagnostischen Praxis – Eine Frage der pädagogischen Zielsetzung? *Empirische Pädagogik*, 28(3), 211–228.